Workgroup Updates

OHDSI Community Call
Oct. 10, 2023 • 11 am ET

Upcoming Community Calls

Date	Topic
Oct. 10	Workgroup Reports
Oct. 17	Symposium Week! Final Logistics + Mad Minutes
Oct. 24	Welcome to OHDSI
Oct. 31	TBA
Nov. 7	Meet The Titans
Nov. 14	Collaborator Showcase Honorees

Three Stages of The Journey

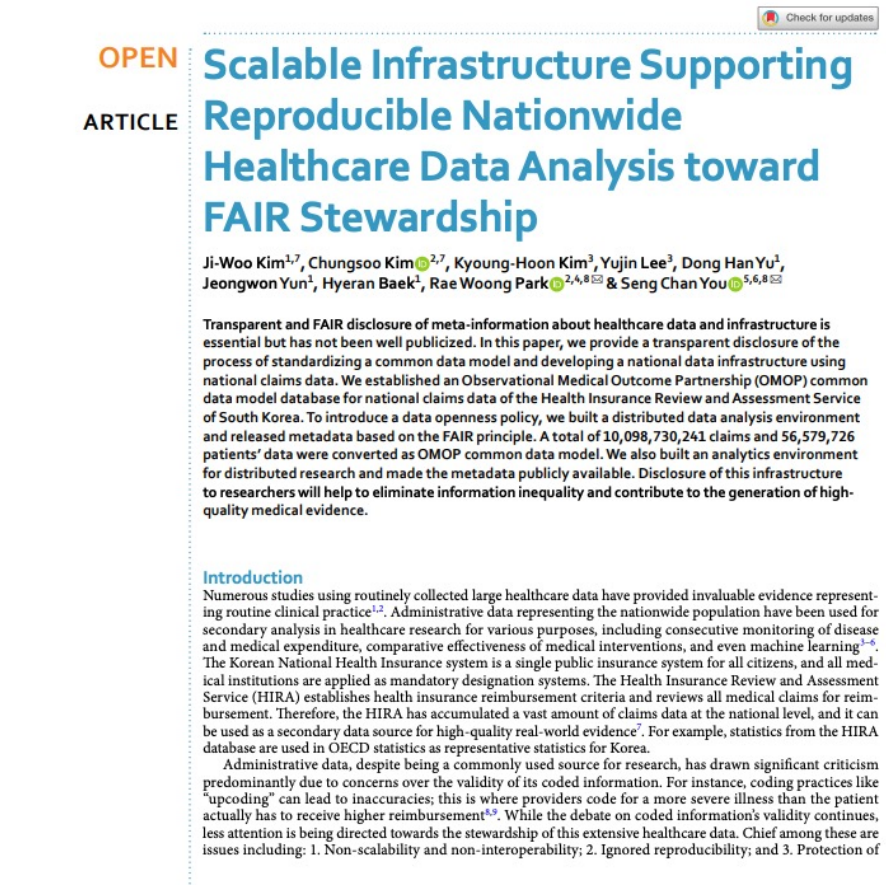
Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Ji-Woo Kim, Chungsoo Kim, Kyoung-Hoon Kim, Yujin Lee, Dong Han Yu, Jeongwon Yun, Hyeran Baek, Rae Woong Park and Seng Chan You** on the publication of **Scalable Infrastructure Supporting Reproducible Nationwide Healthcare Data Analysis toward FAIR Stewardship** in *Scientific Data*.



scientific data

Check for updates

OPEN ARTICLE

Scalable Infrastructure Supporting Reproducible Nationwide Healthcare Data Analysis toward FAIR Stewardship

Ji-Woo Kim[1,7], Chungsoo Kim[2,7], Kyoung-Hoon Kim[3], Yujin Lee[3], Dong Han Yu[1], Jeongwon Yun[1], Hyeran Baek[1], Rae Woong Park[2,4,8] & Seng Chan You[5,6,8]

Transparent and FAIR disclosure of meta-information about healthcare data and infrastructure is essential but has not been well publicized. In this paper, we provide a transparent disclosure of the process of standardizing a common data model and developing a national data infrastructure using national claims data. We established an Observational Medical Outcome Partnership (OMOP) common data model database for national claims data of the Health Insurance Review and Assessment Service of South Korea. To introduce a data openness policy, we built a distributed data analysis environment and released metadata based on the FAIR principle. A total of 10,098,730,241 claims and 56,579,726 patients' data were converted as OMOP common data model. We also built an analytics environment for distributed research and made the metadata publicly available. Disclosure of this infrastructure to researchers will help to eliminate information inequality and contribute to the generation of high-quality medical evidence.

Introduction

Numerous studies using routinely collected large healthcare data have provided invaluable evidence representing routine clinical practice[1,2]. Administrative data representing the nationwide population have been used for secondary analysis in healthcare research for various purposes, including consecutive monitoring of disease and medical expenditure, comparative effectiveness of medical interventions, and even machine learning[3–6]. The Korean National Health Insurance system is a single public insurance system for all citizens, and all medical institutions are applied as mandatory designation systems. The Health Insurance Review and Assessment Service (HIRA) establishes health insurance reimbursement criteria and reviews all medical claims for reimbursement. Therefore, the HIRA has accumulated a vast amount of claims data at the national level, and it can be used as a secondary data source for high-quality real-world evidence[7]. For example, statistics from the HIRA database are used in OECD statistics as representative statistics for Korea.

Administrative data, despite being a commonly used source for research, has drawn significant criticism predominantly due to concerns over the validity of its coded information. For instance, coding practices like "upcoding" can lead to inaccuracies; this is where providers code for a more severe illness than the patient actually has to receive higher reimbursement[8,9]. While the debate on coded information's validity continues, less attention is being directed towards the stewardship of this extensive healthcare data. Chief among these are issues including: 1. Non-scalability and non-interoperability; 2. Ignored reproducibility; and 3. Protection of

OHDSI Shoutouts!

Congratulations to the team of **SeJun Oh, Hyung Joon Joo, Jang Wook Sohn, Sangsoo Park, Jin Su Jang, Jiwon Seong, Kwang Jin Park and Sang Heon Lee** on the publication of **Cloud-based digital healthcare development for precision medical hospital information system** in *Personalized Medicine*.

Future Medicine

JOURNALS BOOKS ABOUT AUTHOR GUIDE SUBMIT

PERSONALIZED MEDICINE, AHEAD OF PRINT | RESEARCH ARTICLE

Cloud-based digital healthcare development for precision medical hospital information system

SeJun Oh, Hyung Joon Joo, Jang Wook Sohn, Sangsoo Park, Jin Su Jang, Jiwon Seong, Kwang Jin Park & Sang Heon Lee

Published Online: 9 Oct 2023 | https://doi.org/10.2217/pme-2023-0074

View Article Tools Share

Abstract

Aim: This study aims to develop a cloud-based digital healthcare system for precision medical hospital information systems (P-HIS). **Methods:** In 2020, international standardization of P-HIS clinical terms and codes was performed. In 2021, South Korea's first tertiary hospital cloud was established and implemented successfully. **Results:** P-HIS was applied at Korea's first tertiary general hospital. Common data model-compatible precision medicine/medical service solutions were developed for medical support. Ultrahigh-quality medical data for precision medicine were acquired and built using big data. Joint global commercialization and dissemination/spreading were achieved using the P-HIS consortium and global common data model-based observational medical outcome partnership network. **Conclusion:** To provide personalized precision medical services in the future, establishing and using big medical data is essential.

Keywords: CDM • digital health care • OMOP • P-HIS • precision medical

OHDSI Shoutouts!

Congrats to our 2023 Titan Award Nominees!

Alexander Davydov • Aniek Markus • Anna Ostropolets • Anthony Sena • Asieh Golozar • Asiyah Lin • Atif Adam • Azza Shoaibi • Can Yin • Carlos Diaz • Center for Surgical Science team • Christian Reich • Christie Quarles • Chungsoo Kim • Cindy Cai • Clair Blacketer • Clark Evans • Craig Sachson • Cynthia Sung Dana Zakrzewski • Danielle Boyce • Davera Gabriel • Debo Wei • Eleanor Davies • Elisse Katzman • Erica Voss • Evan Minty • Frank DeFalco • Geert Byttebier • Georgina Kennedy • Gowtham Rao • Grahame Grieve • Gregory Klebanov • Gyeol Song • Henrik John • Hugo Vernooij • IQVIA OMOP Productized Analytics • Ismail Gogenur • Jack Brewster • James Brash • James Gilbert • Jared Houghtaling • Jasmine Gratton • Jenna Reps • Jiawei Qian • Jiayi (Jessie) Tong • Jing Li • Joel Swerdel • John Gresh • Katherine Duszynski • Katy Sadowski • Kyle Zollo-Venecek • Kyrylo Simonov • LAISDAR Study Team • Lee Evans • Lydia Liu • Manlik Kwong • Marc Suchard • Marc Twagirumukiza • Marcel de Wilde • Masha Khitrun • Marti Catala • Martijn Schuemie • Martin Lavallee • Marty Alvarez • Meghan Pettine • Mengyuan Shang • Michael Matheny • Michelle Hribar • Milou Brand • Montse Camprubi • Nathan Buesgens • Nathan Hall • Nicole Pratt • Nigel Hughes • Nikolai Grewe • OHDSI Vocabulary Team • Oleg Zhuk • Paul Dougall • Paul Nagy • Polina Talapova • Raivo Kolde • Renske Los • Sally Baxter • Sarah Seager • Stephen Town • Tal El-Hay • Thamir Alshammary • Thomas Falconer • Timur Vakhitov • Varvara Savitskaya • Vipina Keloth • Xiaoyu Lin

Winners will be announced during the #OHDSI2023 Closing Talk!

#OHDSISocialShowcase

ohdsi.org/europe2023-showcase

#OHDSISocialShowcase

MONDAY

Implementation of the ARES application to monitor network-wide data quality and mapping coverage for 16 unique OMOP sources across Rwanda

(Jared Houghtaling, Emma Gesquiere, Lars Halvorsen, Marc Twagirumukiza, and Charles Ruranga)

As we finalize the LAISDAR project, we are set to begin the **next chapter of OMOP-based federated networks** both in Rwanda and around the African continent; **Ares**, along with other powerful **open-source OHDSI tooling**, will be central in those efforts

Title: Implementation of the ARES application to monitor network-wide data quality and mapping coverage for 16 unique OMOP sources across Rwanda

BACKGROUND: Leveraging Artificial Intelligence and Data Science Techniques in Harmonizing, Accessing and Analysing SARS-COV-2/COVID-19 Data in Rwanda, or LAISDAR, aims to establish a nation-wide federated data network based on the Observational Medical Outcomes Partnership (OMOP) common data model (CDM) [1, 2]. The project was initially intended to support research on COVID-19 but given the quantity and quality of electronic health record (EHR) data available at the various participating hospitals, the scope has since widened to other relevant communicable and noncommunicable disease areas. Most Rwandan hospitals have implemented one of two EHR systems for storing electronic medical data: openClinic GA and openMRS [3]. A first step in the project was to define structural and semantic mappings and logic to transform the two source systems to the OMOP CDM. We have since developed an Extract-Transform-Load (ETL) process that can transform data in either source EHR system to the target format and that uses a variable switch to select EHR-specific transformations and mappings. Currently, the network comprises 16 OMOP instances; 14 different hospitals have EHR data in OMOP format, with a combined total of more than 3,5M (approximately 25% of the national population) individuals represented. Additionally, we have transformed national data related to test results for COVID-19 into an independent OMOP dataset, and we have also transformed COVID-19-related survey data from a 10'000+ participant survey conducted in early 2022 into a separate OMOP instance. Tracking data quality, mapping coverage, and transformation versions across a network of this size is nontrivial; for this task, we have employed a set of new OHDSI tools, Ares [4] and AresIndexer [5], which when combined compile aggregate statistics and information about a set of OMOP data sources and releases for validation and exploration. In this work, we present our experiences using Ares, with the intention of (1) highlighting the power and ease-of-use of the tool, and (2) motivating others facing similar multi-OMOP-source challenges to implement the tool as a plug-and-play solution.

Figure 1: Services deployed (via Docker) on each hospital's Mac Mini machine, and directional connections between the Mac Mini and central server, as well as between the Mac Mini and the local EHR system

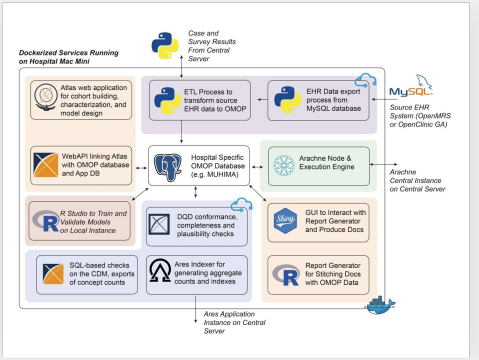
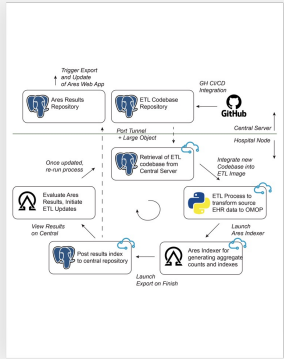


Figure 2: Feedback loop for quality control, relying on SimpleMDM for scripting automation and Ares for network-wide overviews and mapping prioritization



LIMITATIONS AND DISCUSSION: Ares has been tremendously useful in compiling a network-wide overview of pain points, data gaps, and deployment status across this network. Importantly, the AresIndexer relies on common dependencies (e.g. Achilles) used by other OHDSI tooling, so integrating it into the deployment workflow is both simple and computationally efficient. In this case, we created two separate docker images (Ares web application & AresIndexer) that can be deployed quickly and easily on different operating systems and architectures. We plan to make these images publicly available in the coming weeks. Used in combination with a remote orchestration tool like SimpleMDM, Ares completes a powerful feedback loop for ensuring data quality and enables a detailed overview that is critical for defining network-wide studies. The most important functionalities Ares has provided in this project are: (1) a dynamic benchmark that motivates data managers to improve their local data quality and to participate actively in the network, (2) an overview of the diverse data quality issues that stem from subtle differences (e.g. date handling, local mappings) between EHR system configurations, and (3) a detailed tool for evaluating feasibility and potential impact of future federated studies. Given that the Ares tool is still in beta testing, we look forward to updated functionality and capabilities in future releases, and we plan to take an active role in its development moving forward.

OHDSI

LAISDAR

Jared Houghtaling[a], Emma Gesquiere[a], Lars Halvorsen[a], Marc Twagirumukiza[b], and Charles Ruranga[c]

AFFILIATIONS
a external health NV (Kortrijk, Belgium)
b University of Gent (Ghent, Belgium)
c University of Rwanda (Kigali, Rwanda)

REFERENCES
[1] https://hic.gov.rw/laisdar/
[2] Nshimyve, A., Ruranga, C., Musanabaganwa, C. et al. 'Leveraging artificial intelligence and data science techniques in harmonizing, sharing, accessing and analysing SARS-COV-2/COVID-19 data in Rwanda (LAISDAR Project): study design and rationale.' BMC Medical Informatics and Decision Making 22, 214 (2022).
[3] Akanbi, Maxwell O., et al. "Use of electronic health records in sub-Saharan Africa: progress and challenges." Journal of Medicine in the Tropics 14.1 (2012): 1.
[4] Ohdsi. (n.d.). GitHub - OHDSI/Ares: A Real-world Insights on Systems. GitHub. https://github.com/OHDSI/Ares
[5] Ohdsi. (n.d.). GitHub - OHDSI/AresIndexer: R package that creates the index and relevant files for an Ares deployment. GitHub. https://github.com/OHDSI/AresIndexer

edence Health

#OHDSISocialShowcase

TUESDAY

Conceptual architecture for the Digital Oncology Network for Europe - an OMOP based European federated, automated cancer care quality ecosystem

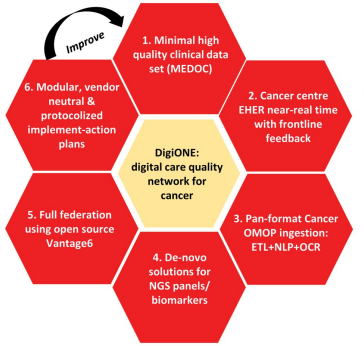
(Piers Mahon, Olivier Bouissou, Ismini Chatzitheofilou, Gennaro Ciliberto, Marco Denti, Xosé Fernández, Dennis Kadioglu, Stelios Theophanous, Joëlle Thonnard, Alberto Traverso)

DigiONE aims to normalise 36 data concepts across 6 European countries to improve the management of cancer

*Conceptual architecture for **Digi**tal **O**ncology **N**etwork for **E**urope (DigiONE) - an OMOP based European federated, automated cancer care quality ecosystem*

Background: Digital methods could provide nearly **real-time information on clinical practice** and outcomes with minimal to no manual retyping of data. This can allow analysis of real-time **medical guideline compliance**. However, there are **many challenges** such as EHR system heterogeneity, low data completeness in key clinical phenotype data, biomarker data often in PDFs, complicated definitions such as line of therapy, hospital IT capacity and privacy/GDPR. DigiONE tackles these challenges through an innovative, integrated **multi-modal OMOP NLP solution with federation**.

Result 1: The key features of DigiONE

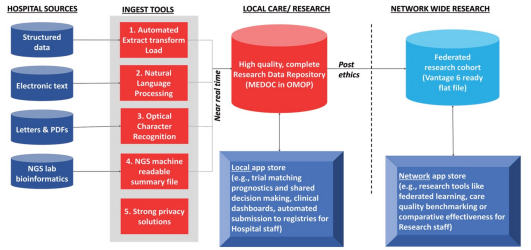


1: **Minimal Essential Description Of Cancer** (MEDOC): 36 data concepts chosen by multi-country clinical consensus
2: **Near-real time frontline feedback loops** to improve data
3: **Pan-format cancer data ingestion**. Not just ETL also NLP, OCR
4: **GDPR recital 34 privacy conserving solutions for NGS**
5: **Federation with open source Vantage6** to allow statistical analysis equivalent to centralised data, but without data pooling
6: **Modular, protocolized implementation plans** to solve for varying data normalisation skills in most hospitals. DigiONE places emphasis on **care quality** as the primary use case for data.
7: **All in open standards and vendor agnostic**

Result 2: Ingestion tooling and architecture for a cancer centre cover ETL, NLP, OCR

Methods

- DIGICORE invited care quality focused hospitals to apply for funding in a **two-step implementation process**.
- Entries were judged by an independent expert committee including patient representation.
- The prototype must achieve high routine data quality on the target dataset MEDOC with appropriate privacy management under GDPR.
- **Six hospitals** were awarded funding.



Piers Mahon, Olivier Bouissou, Ismini Chatzitheofilou, Gennaro Ciliberto, Marco Denti, Xosé Fernández, Dennis Kadioglu, Stelios Theophanous, Joëlle Thonnard, Alberto Traverso

OHDSI

#OHDSISocialShowcase

WEDNESDAY

Paving the way to estimate dose in OMOP CDM for Drug Utilisation Studies in DARWIN EU

(Theresa Burkard, Kim Lopez-Güell, Artem Gorbachev, Annika M Jödicke, Nuria Mercadé-Besora, Talita Duarte-Salles, Dani Prieto-Alhambra, Christian Reich, Marti Catala)

This approach shall **maximize the completeness and reliability of dose estimations in OMOP CDM** through identification of relevant units and patterns among the drug strength table.

Paving the way to estimate dose in OMOP CDM for Drug Utilisation Studies in DARWIN EU®

Background: Currently, there is no standardized comprehensive method to infer daily dose of drugs from OMOP CDM drug exposure records.

Results: The review of the patterns among the drug strength table identified the relevant units for dose estimation (Table 1).

Table 1. Relevant drug strength patterns for dose estimation

amount	amount_unit	numerator	numerator_unit	denominator	denominator_unit
numeric	international unit	NA	NA	NA	NA
numeric	microgram	NA	NA	NA	NA
numeric	milliequivalent	NA	NA	NA	NA
numeric	milligram	NA	NA	NA	NA
numeric	milliliter	NA	NA	NA	NA
NA	NA	numeric	international unit	numeric	milligram
NA	NA	numeric	international unit	NA	milligram
NA	NA	numeric	international unit	numeric	milliliter
NA	NA	numeric	international unit	NA	milliliter
NA	NA	numeric	mega international unit	NA	milliliter
NA	NA	numeric	microgram	numeric	hour
NA	NA	numeric	microgram	NA	hour
NA	NA	numeric	milliequivalent	NA	milligram
NA	NA	numeric	milliequivalent	numeric	milliliter
NA	NA	numeric	milliequivalent	NA	milliliter
NA	NA	numeric	milligram	numeric	Actuation
NA	NA	numeric	milligram	NA	Actuation
NA	NA	numeric	milligram	numeric	hour
NA	NA	numeric	milligram	NA	hour
NA	NA	numeric	milligram	numeric	liter
NA	NA	numeric	milligram	NA	liter
NA	NA	numeric	milligram	numeric	milligram
NA	NA	numeric	milligram	NA	milligram
NA	NA	numeric	milligram	numeric	milliliter
NA	NA	numeric	milligram	NA	milliliter
NA	NA	numeric	milligram	numeric	square centimeter
NA	NA	numeric	milligram	NA	square centimeter
NA	NA	numeric	milliliter	numeric	milligram
NA	NA	numeric	milliliter	NA	milligram
NA	NA	numeric	milliliter	numeric	milliliter
NA	NA	numeric	milliliter	NA	milliliter

Among the concentration patterns (i.e. combinations of numerator and denominator units), only few patterns are used according to the drug exposure table. Most frequently used patterns have missing denominator values (Table 2).

Table 2. Counts of unique drug concept ids (in combination with quantity) per concentration pattern that were used in CPRD Gold and CPRD AURUM

pattern name	counts in CPRD GOLD	counts in CPRD AURUM
international unit per milligram	0	0
international unit per milligram missing denominator	0	0
international unit per milliliter	0	0
international unit per milliliter missing denominator	0	0
mega international unit per milliliter	0	0
mega international unit per milliliter missing denominator	0	0
microgram per hour	0	0
microgram per hour missing denominator	0	0
milliequivalent per milligram	0	0
milliequivalent per milligram missing denominator	0	0
milliequivalent per milliliter	0	0
milliequivalent per milliliter missing denominator	0	0
milligram per actuation	0	0
milligram per actuation missing denominator	**8959**	**5369**
milligram per hour	4457	2220
milligram per hour missing denominator	**1317**	**1070**
milligram per liter	0	0
milligram per liter missing denominator	0	0
milligram per milligram	2180	1200
milligram per milligram missing denominator	**34058**	**15167**
milligram per milliliter	23316	10930
milligram per milliliter missing denominator	**83959**	**24338**
milligram per square centimeter	0	0
milligram per square centimeter missing denominator	0	97
milliliter per milligram	0	0
milliliter per milligram missing denominator	0	0
milliter per milliliter	0	0
milliter per milliliter missing denominator	**105**	**0**

Methods: We created patterns from the drug strength table of OMOP CDM by grouping it by amount_value, amount_unit, numerator_value, numerator_unit, denominator_value, and denominator_unit, to disentangle the units which are crucial for dose estimations. The resulting patterns were independently assessed for relevant units/patterns by two pharmacists (AJ, TB) and a medical doctor (AG). In a consensus meeting with CR, relevant units/patterns and unit standardizations were decided. Dose estimation in concentration patterns (i.e. combinations of numerator and denominator units) with missing denominator values may be difficult, therefore, we assessed the frequency of used unique drug concept ids (in combination with quantity) per pattern in the drug exposure table which would give us a priority list for dose formula development. All checks were performed in CPRD GOLD and CPRD AURUM.

Limitation: This work was only carried out in CPRD GOLD and CPRD AURUM and shall be continued in more databases including more countries and settings (claims, hospital data).

Theresa Burkard[1], Kim Lopez-Güell[1], Artem Gorbachev[2], Annika M Jödicke[1], Nuria Mercadé-Besora[3], Talita Duarte-Salles[3,4], Maria de Ridder[4], Mees Mosseveld[4], Daniel Prieto-Alhambra[1,4], Christian Reich[4], Martí Català[1]

[1] Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK
[2] Odysseus Inc
[3] Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol)
[4] Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands

#OHDSISocialShowcase

THURSDAY

Pre-coordination of structural and semantic conventions to capture breast cancer indicators in the context of the European University Hospital Alliance (EUHA)

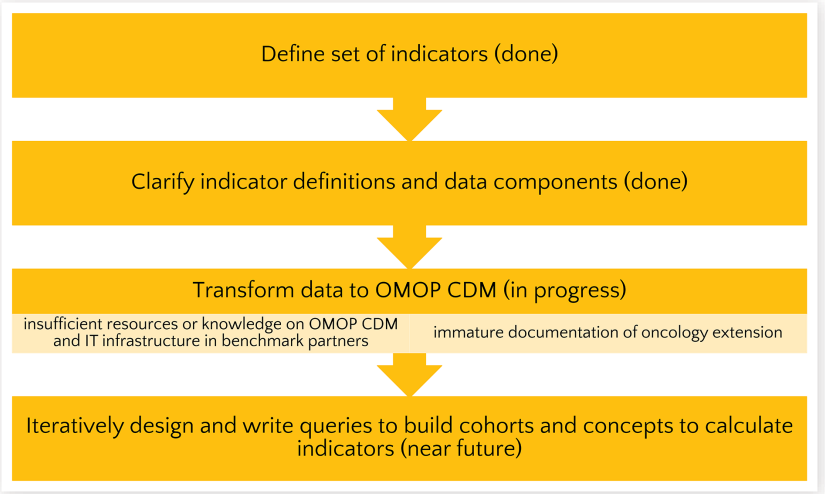
(Tijs Delabastita, Pieter Maertens, Jared Houghtaling, Pieter Stijnen, Guy Vanden Boer, Johan Van Eldere)

More partners participating in EHDEN Data Partner Calls might speed up breast cancer quality of care benchmark between European university hospitals.

Pre-coordination of structural and semantic conventions to capture breast cancer indicators in the context of the European University Hospital Alliance

Background: UZ Leuven coordinates a study comparing breast cancer quality indicators between partners from the European University Hospital Alliance. The study uses data transformed to the OMOP CDM. The diversity in OMOP data source maturity across the partners is considerable.

Figure 1. Schematic overview of the pre-coordination process and main corresponding challenges



UZ LEUVEN OHDSI edence Health

Tijs Delabastita, Pieter Maertens, Jared Houghtaling, Pieter Stijnen, Guy Vanden Boer, Johan Van Eldere

#OHDSISocialShowcase

FRIDAY

Toward generalizable clinical prediction models: a large-scale investigation using stacking ensembles

(Cynthia Yang, Egill A. Fridgeirsson, Jan A. Kors, Jenna M. Reps, Peter R. Rijnbeek, Jenna Wong, Ross D. Williams)

Stacking can improve performance over using the individual base learners alone.

Toward generalizable clinical prediction models: a large-scale investigation using stacking ensembles

Background: Stacking ensembles exploit the heterogeneous properties of multiple models (base learners) to obtain better predictions from the ensemble than from an individual base learner. We performed a large-scale investigation of stacking combining multiple base learners of different algorithm types trained within a single observational health database. We investigated whether stacking (using logistic regression as meta-learner) improves internal and external validation performance across 4 databases compared to 6 individual base learners, using 21 outcomes within a target population of people with pharmaceutically treated depression.

Figure 1: AUC differences summarized across all outcomes on internal and external validation (each column represents a development database, and each row represents a validation database).

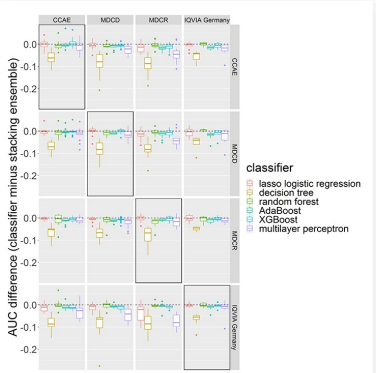
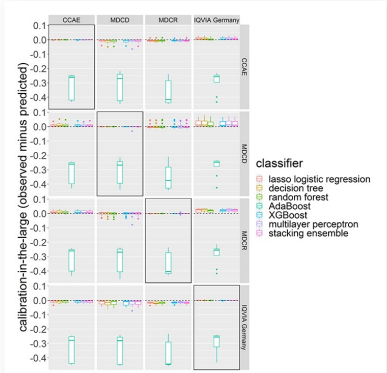


Figure 2: Calibration-in-the-large summarized across all outcomes on internal and external validation (each column represents a development database, and each row represents a validation database).



Findings: The stacking ensembles consistently performed as well as or better than the individual base learners on internal and external validation. Stacking can improve model generalizability and can be a good way to define an upper bound of performance for a particular prediction task.

OHDSI

Cynthia Yang, Egill A. Fridgeirsson, Jan A. Kors, Jenna M. Reps, Peter R. Rijnbeek, Jenna Wong, Ross D. Williams

Erasmus MC Rotterdam
Health Data Science

OHDSI Shoutouts!

Any shoutouts from the community? Please share and help promote and celebrate OHDSI work!

Do you have anything you want to share? Please send to sachson@ohdsi.org so we can highlight during this call and on our social channels.
Let's work together to promote the collaborative work happening in OHDSI!

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

Upcoming Workgroup Calls

Date	Time (ET)	Meeting
Tuesday	3 pm	OMOP CDM Oncology Outreach/Research Subgroup
Wednesday	9 am	Patient-Level Prediction
Wednesday	12 pm	Health Equity
Wednesday	2 on	Natural Language Processing
Thursday	8 am	India Chapter
Thursday	9:30 am	Data Network Quality
Thursday	7 pm	Dentistry
Friday	9 am	GIS – Geographic Information System Development
Friday	9 am	Phenotype Development & Evaluation
Friday	1 pm	Clinical Trials
Friday	11 pm	China Chapter
Monday	9 am	Vaccine Vocabulary
Monday	10 am	Africa Chapter
Monday	11 am	Data Bricks User Group
Monday	6 pm	OMOP & FHIR

October Newsletter Is Available

OHDSI

The Journey Newsletter (October 2023)

It is officially Symposium Month! The 2023 OHDSI Global Symposium will be held Oct. 20-22 at the Hilton East Brunswick Hotel & Executive Meeting Center in East Brunswick, N.J., and there is still time to register. This newsletter takes a closer look at the symposium weekend, including the three-day collaborator showcase, while also reflecting on the growth of the DARWIN EU® Initiative. We also celebrate the 90+ Titan Award nominees! **#JoinTheJourney**

OHDSI Videocast: Symposium Showcase, DARWIN EU®, Titans & more

In the latest On The Journey video, Patrick Ryan and Craig Sachson take a deeper look at the expanded Collaborator Showcase for the 2023 OHDSI Global Symposium main conference. They discuss the 90+ Titan Award nominees who were nominated by members of the community, and they reflect on two European initiatives that were highlighted during September community calls, DARWIN EU® and the 11th Revision of the ENCePP Guide on Methodological Standards in Pharmacoepidemiology. *(if video does not appear, click 'view this email in your browser')*

Community Updates

Where Have We Been?

- The Sept. 5 OHDSI Community Call provided a look at the DARWIN EU® Progress and Roadmap. DARWIN EU® is an initiative that works to deliver real-world evidence from across Europe on diseases, populations and the uses and performance of medicines. Members of the leadership team provided an update on both its progress and goals.
- ENCePP editor Catherine Cohet and seven authors of the 11th Revision of the ENCePP Guide on Methodological Standards in Pharmacoepidemiology led a discussion on the purpose and audience of the guide, some of the specific focuses within this revision, and how it can benefit the research mission of the OHDSI community.

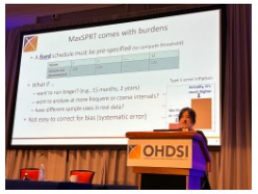
Where Are We Now?

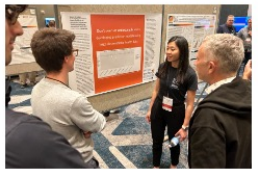
- After months of preparation, the OHDSI Global Symposium is less than three weeks away. The Global Symposium will be held Oct. 20-22 at the Hilton East Brunswick Hotel & Executive Meeting Center in East Brunswick, N.J. The full weekend agenda is now available, and it includes the Friday conference agenda, descriptions on the various weekend activities, and all 137 posters and 24 software demos that will be presented during the collaborator showcase.
- Congratulations to the 93 individuals/teams who were nominated for a 2023 Titan Award. Each year, Titan Awards are presented to members of the community who have made valuable contributions to OHDSI's mission, vision and values. The recipients of the seven Titan Awards will be announced during the closing talk on Friday, Oct. 20 at the Global Symposium.

Where Are We Going?

- We are going to East Brunswick, of course, and we hope to see you there. We are still accepting registrations for the Global Symposium, but don't wait too long to secure your spot at the highlight event of the OHDSI year! Use this link to register for #OHDSI2023!
- The first two weeks of OHDSI community calls this month (Tuesdays, 11 am ET) will feature updates from our various workgroups, and they will help set the foundation for our symposium weekend activities. Community calls are open to everybody, and they are a great way to stay updated on everything happening within the community. The meeting link is available within our Community Calls page.

#OHDSI2023 Month is Here! Check Out The Full Weekend Agenda For The Oct. 20-22 Global Symposium



The agenda for the 2023 Global Symposium main conference is now available and it highlights the most diverse agenda in our event history. It lists the full schedule for all three days, descriptions on the various weekend activities, and all 137 posters and 24 software demos that will be presented during the collaborator showcase.

Register for the 2023 Global Symposium

2023 Global Symposium Weekend Agenda

Collaborator Showcase Posters

Collaborator Showcase Software Demos

Collaborator Showcase Lightning Talks

September Publications

Przysucha M, Hüsers J, Liberman D, Kersten O, Schlüter A, Fraas S, Busch D, Moelleken M, Erfurt-Berge C, Dissemond J, Hübner U. Design and Implementation of an ETL-Process to Transfer Wound-Related Data into a Standardized Common Data Model. Stud Health Technol Inform. 2023 Sep 12;307:258-266. doi: 10.3233/SHTI230723. PMID: 37697861.

de Groot R, Püttmann DP, Fleuren LM, Thoral PJ, Elbers PWG, de Keizer NF, Cornet R; Dutch ICU Data Sharing Against COVID-19 Collaborators. Determining and assessing characteristics of data element names impacting the performance of annotation using Usagi. Int J Med Inform. 2023 Aug 29;178:105200. doi: 10.1016/j.ijmedinf.2023.105200. Epub ahead of print. PMID: 37703800.

Lee S, Shin H, Choe S, Kang MG, Kim SH, Kang DY, Kim JH. MetaLAB-HOI: Template standardization of health outcomes enable massive and accurate detection of adverse drug reactions from electronic health records. Pharmacoepidemiol Drug Saf. 2023 Sep 14. doi: 10.1002/pds.5694. Epub ahead of print. PMID: 37710363.

Raventós B, Fernández-Bertolín S, Aragón M, Voss EA, Blacketer C, Méndez-Boo L, Recalde M, Roel E, Pistillo A, Reyes C, van Sandijk S, Halvorsen L, Rijnbeek PR, Burn E, Duarte-Salles T. Transforming the Information System for Research in Primary Care (SIDIAP) in Catalonia to the OMOP Common Data Model and Its Use for COVID-19 Research. Clin Epidemiol. 2023 Sep 13;15:969-986. doi: 10.2147/CLEP.S419481. PMID: 37724311; PMCID: PMC10505380.

Moya A, Oeste CL, Beles M, Verstreken S, Dierckx R, Heggermont W, Bartunek J, Bogaerts E, Masuy I, Hens D, Bertolone D, Vanderheyden M. Detection of transthyretin amyloid cardiomyopathy by automated data extraction from electronic health records. ESC Heart Fail. 2023 Sep 19. doi: 10.1002/ehf2.14517. Epub ahead of print. PMID: 37726928.

Matcham F, Simblett SK, Leightley D, Dalby M, Siddi S, Haro JM, Lamers F, Penninx BWHJ, Bruce S, Nica R, Zormpas S, Gilpin G, White KM, Oetzmann C, Annas P, Brasen JC, Narayan VA, Hotopf M, Wykes T; RADAR-CNS consortium. The association between persistent cognitive difficulties and depression and functional outcomes in people with major depressive disorder. Psychol Med. 2023 Oct;53(13):6334-6344. doi: 10.1017/S0033291722003671.

mailchi.mp/ohdsi/october2023

October Newsletter Is Available

mailchi.mp/ohdsi/october2023

Global Symposium

Oct. 20-22 • East Brunswick, NJ, USA

Hilton East Brunswick Hotel & Executive Meeting Center

bit.ly/OHDSI2023Registration

Global Symposium Conference Agenda

Agenda • Friday, Oct. 20

Time	Topic
7:30 - 8:30 am East Brunswick Room + Grand Ballroom Foyer	Symposium Registration, Lite Breakfast Buffet, All-Day Exhibits * First-timers can meet for a quick orientation session at 7:45 am in Piscataway/Woodbridge (will conclude before the start of the first talk)
8:30 - 9:30 am Grand Ballroom	State of the Community OHDSI: Where have we been? Where are we going? George Hripcsak, Columbia Univ. Community Highlights: • OMOP CDM users and the OHDSI data network Clair Blacketer, Johnson & Johnson • OHDSI standardized vocabularies Alexander Davydov, Odysseus Data Services • OHDSI's open-source community Katy Sadowski, Boehringer Ingelheim • OHDSI Europe 2024 Peter Rijnbeek, Erasmus MC • OHDSI Asia-Pacific 2024 Mengling Feng, National Univ. of Singapore
9:30 - 10:30 am Grand Ballroom	OHDSI Community Networking Moderators: • Faaizah Arshad, Univ. of California-Los Angeles • Cynthia Sung, Duke-NUS Medical School
10:30 am - 12:00 pm Grand Ballroom	Plenary: Improving the reliability and scale of case validation Presenters: • Patrick Ryan, Johnson & Johnson, Columbia Univ. • Anna Ostropolets, Odysseus Data Services • Martijn Schuemie, Johnson & Johnson, Univ. of California-Los Angeles
12:00 pm - 1:00 pm Grand Ballroom Foyer	Buffet Lunch

All events take place at the Grand Ballroom Level • Exhibits will be available throughout the day

Time	Topic
1:00 pm - 2:00 pm Grand Ballroom	Panel: Lessons learned from OHDSI network studies Presenters: • Insights from LEGEND-T2DM Marc Suchard, Univ. of California-Los Angeles • Intravitreal anti-VEGF and risk of kidney failure: A Sisyphus Challenge Study Cindy X Cai, Johns Hopkins Univ. • Fluoroquinolones and the risk of aortic aneurysm: A Sisyphus Challenge study Seng Chan You, Yonsei Univ. • Lessons learned applying the Strategus framework across the OHDSI network Anthony Sena, Johnson & Johnson Moderator: Sarah Seager, IQVIA
2:00 pm - 2:45 pm Grand Ballroom	Collaborator Showcase, Lightning Talk Session #1: Data Standards and Methods Research • Mapping of Critical Care EHR Flowsheet data to the OMOP CDM via SSSOM Polina Talapova, SciForce • Paving the way to estimate daily dose in OMOP CDM for Drug Utilisation Studies in DARWIN EU® Theresa Burkard, Univ. of Oxford • Generating Synthetic Electronic Health Records in OMOP using GPT Chao Pang, Columbia Univ. • Comparing concepts extracted from clinical Dutch text to conditions in the structured data Tom Seinen, Erasmus MC • Finding a constrained number of predictor phenotypes for multiple outcome prediction Jenna Reps, Johnson & Johnson Moderator: Davera Gabriel, Johns Hopkins University
2:45 - 3:30 pm Grand Ballroom	Collaborator Showcase, Poster / Demo Session #1 Poster walk leads: • Data standards: Mui Van Zandt, IQVIA • Methods research: Christophe Lambert, Univ. of New Mexico • Open-source development: Paul Nagy, Johns Hopkins Univ. • Clinical applications: Kristin Kostka, Northeastern University

Time	Topic
3:30 pm - 4:15 pm Grand Ballroom	Collaborator Showcase, Lightning Talk Session #2: Methods Research and Clinical Applications • Synthesizing Evidence for Rare Events: a Novel Zero-Inflated Bivariate Model to Integrate Studies with Double-Zero Outcomes Lu Li, Univ. of Pennsylvania • Active Safety Surveillance Using Real-world Evidence (ASSURE): An application of the Strategus package Kevin Haynes, Johnson & Johnson • Patient's outcomes after endoscopic retrograde cholangiopancreatography (ERCP) using reprocessed duodenoscope accessories: a descriptive study using real-world data Jessica Maruyama, Precision Data • Does COVID-19 Increase Racial/Ethnic Differences in Prevalence of Post-acute Sequelae of SARS-CoV-2 infection (PASC) in Children and Adolescents? An EHR-Based Cohort from the RECOVER Program Bingyu Zhang, Univ. of Pennsylvania • Eye Care and Vision Research Workgroup: First Year Update Michelle Hribar, National Institutes of Health – National Eye Institute Moderator: Atif Adam, IQVIA
4:15 - 5:00 pm Grand Ballroom	Collaborator Showcase, Poster / Demo Session #2 Poster walk leads: • Data standards: Melanie Philofsky, Odysseus Data Services • Methods research: Andrew Williams, Tufts Univ. • Open-source development: Nsikak Akpakpan, Accenture • Clinical applications: Hanieh Razzaghi, Childrens Hospital of Pennsylvania
5:00 pm - 6:00 pm Grand Ballroom	Closing session: Scaling community, scaling collaboration • Titan Awards • Group Photo Presenter Patrick Ryan, Johnson & Johnson, Columbia Univ.
6:00 pm - 7:00 pm East Brunswick Room Grand Ballroom Foyer	Networking Reception and Exhibits
7:00 pm - 8:00 pm Grand Ballroom	OHDSI Got Talent!

bit.ly/OHDSI2023Agenda

Register

Global Symposium Conference Agenda

Agenda • Saturday, Oct. 21

Time	Topic
7:00 - 8:00 am Grand Ballroom Foyer	Lite Breakfast Buffet, All-Day Exhibits
8:00 am - 12:00 pm Various rooms	Introduction to OHDSI Tutorial Common Data Model/Network Data Quality WG Meeting Health Analytics Data-to-Evidence Suite (HADES) Hackathon Health EquityWG Meeting Medical Imaging WG Meeting Natural Language Processing WG Meeting OHDSI Industry WG Kickoff Meeting Oncology WG Meeting Phenotype Development & Evaluation WG Meeting Pregnancy and Reproductive Health Group (PRHeG) WG Meeting
12:00 - 1:00 pm Ballroom Foyer/ Ballroom	Lunch Buffet, Collaborator Showcase, All-Day Exhibits
1:00 pm - 5:00 pm Grand Ballroom	HowOften Large-Scale Characterization Workshop
5:00 pm	Free Time

Agenda • Sunday, Oct. 22

Time	Topic
7:00 - 8:00 am Grand Ballroom Foyer	Lite Breakfast Buffet, All-Day Exhibits
8:00 am - 12:00 pm Grand Ballroom/ Room TBA	HowOften Large-Scale Characterization Workshop HL7 FHIR-OMOP Connectathon
12:00 - 1:00 pm Ballroom Foyer/ Ballroom	Lunch Buffet, Collaborator Showcase, All-Day Exhibits
1:00 pm - 5:00 pm Various Rooms	Africa Chapter Workshop Eye Care & Vision Research WG Meeting Health Analytics Data-to-Evidence Suite (HADES) Hackathon Healthcare Systems Interest Group (HSIG) WG Meeting HL7 FHIR-OMOP Connectathon ISPE RWE for Pharmacovigilance Medical Devices WG Meeting Psychiatry WG Meeting Vocabulary WG Meeting Latin America WG Meeting
5:00 pm	Symposium Closing

bit.ly/OHDSI2023-Agenda

Register

Welcome, 1st-Time Attendees!

All OHDSI first-time attendees are welcome to attend an orientation on Friday at 7:45 am within the Woodbridge/ Piscataway room. **Paul Nagy**, a 2022 Titan honoree for community leadership, will lead this session.

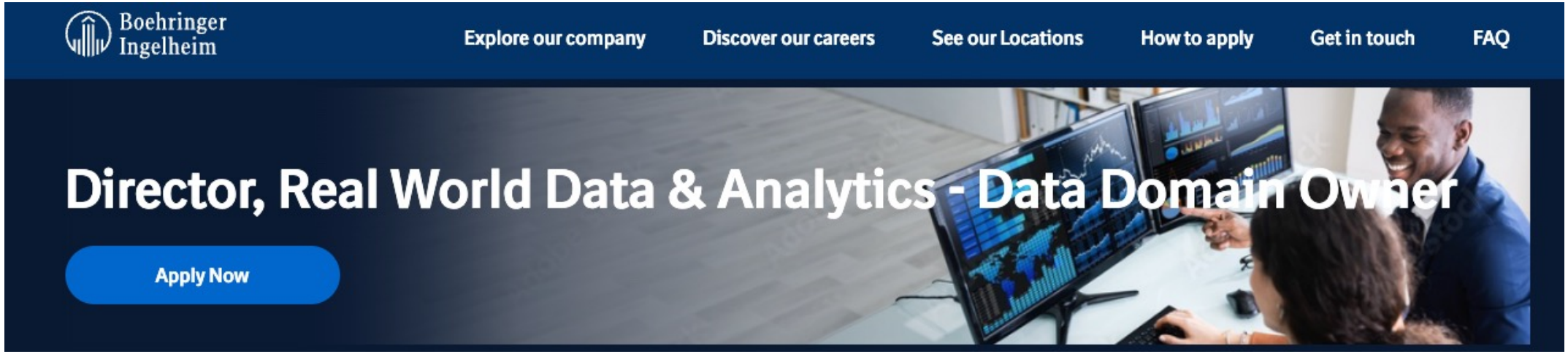
OHDSI

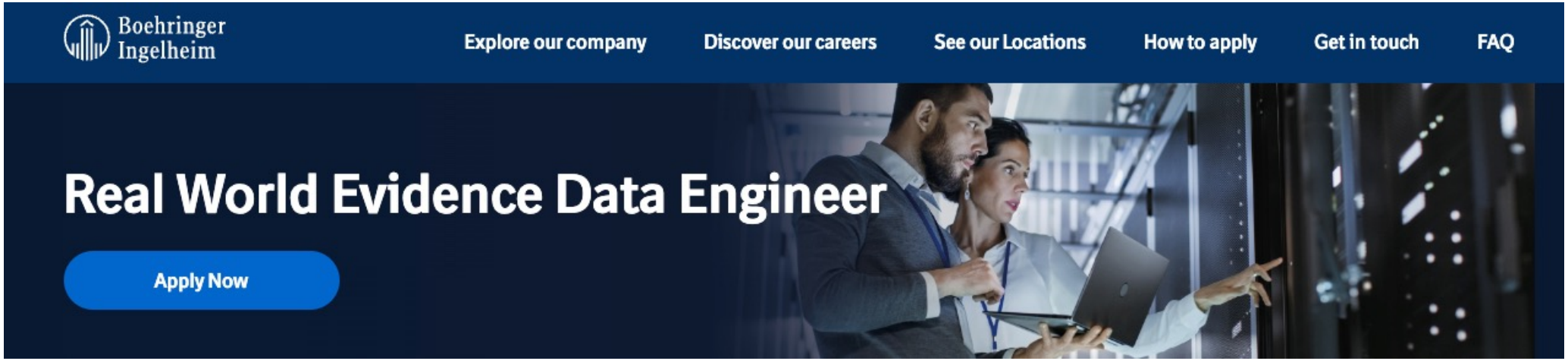
1st Time Attendees

ohdsi.org/ohdsi2023

Register

Openings at Boehringer Ingelheim





Opening: Postdoctoral Associate/Data Analyst

Job Announcement: Postdoctoral Associate/Data Analyst - LEGEND Hypertension Project

Position: Postdoctoral Associate/Data Analyst

Organization: Yale University, School of Medicine

Location: 195 Church Street, 5th floor, New Haven, CT, 06510

Application Deadline: Rolling basis

Job Description:

We are seeking a talented and dedicated Postdoctoral Associate/Data Analyst to join our dynamic team. In this role, you will play a pivotal part in advancing our mission of improving health outcomes through data-driven research. You will have the opportunity to work with diverse healthcare datasets, develop innovative analytical methods, and collaborate with experts in the field.

The Postdoctoral Associate/Data Analyst should possess significant experience in R and Rstudio, with specific expertise in database management using PostgreSQL—critical requirements within the OHDSI network. Your responsibilities will include assisting the Principal Investigator (Dr. Yuan Lu from Yale University) and Co-Investigator (Drs. Marc Suchard from UCLA) in creating the analytic tool stack and performing related analyses.

Key Responsibilities:

- Collaborate with multidisciplinary teams to design and execute data analysis projects.
- Develop and implement statistical and machine learning models for healthcare data.
- Perform data extraction and preprocessing tasks to prepare datasets for analysis.
- Conduct exploratory data analysis and visualization to extract insights from healthcare data.
- Assist in the development and maintenance of OHDSI's open-source tools and resources.
- Communicate findings and insights through reports, presentations, and publications.
- Stay up-to-date with the latest advancements in data science and healthcare informatics.

Email: y.lu@yale.edu

Where Are We Going?

Any other announcements of upcoming work, events, deadlines, etc?

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI's Open Source Working Group

Adam Black, Paul Nagy

What is the Open Source Working Group

- The OHDSI Open Source Community exists to promote the health and sustainability of the OHDSI open source software ecosystem. This entails building open-source governance and engaging new contributors.
- 271 Members of Teams Working Group

	Last 12 years	Last 12 months
Repositories	262	134
Contributors	681	210
Commits	47,672	6,160
Organizations Contributing	31	20

Open Source Working Group

- OKR 1: Improve process to engage and train casual contributors to become active contributors/maintainers
  - Expand leadership team. Katy Sadowski, Clark Evans, Nate Buesgens, Dan Smith, Anthony Sena.
  - 2023 Kheiron Cohort. 20 new contributors with 1:1 mentors.
  - 2023 OHDSI DevCon

OKR 2: Formation of a Technical Advisory Board (TAB).

The mission of the OHDSI Technical Advisory Board is to ensure the stability, security, supportability, and sustainability of OHDSI open source projects.

TAB Lead:

Lee Evans

2023 Achievements:

- 14 representatives from across the OHDSI ecosystem joined the TAB and drafted a charter
- Kicked off work to:
  - Align on and implement **standards for database platform support** (including shared testing infrastructure)
  - Develop technical and process solutions for **coordinated, stable, and secure OHDSI software releases**

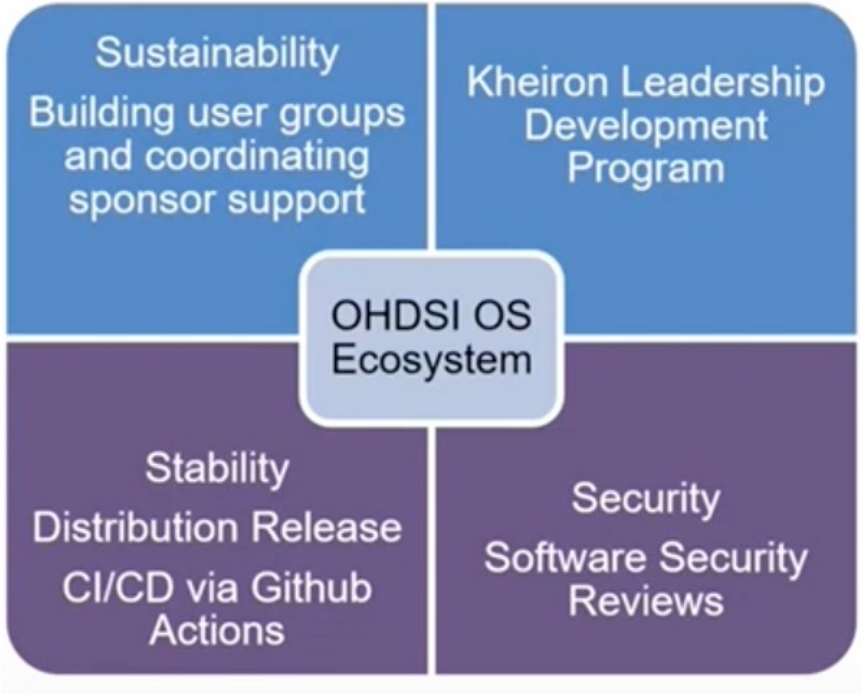


Image source: Paul Nagy

Join the Journey!

Health Analytics Data-to-Evidence Suite (HADES) Hackathon

- Saturday 8:00am-12:00pm and Sunday 1:00pm-5:00pm
- Participants will work on the HADES codebase with support from several HADES maintainers. Participants can work in groups, and we welcome both new and experienced contributors to join
- Target audience: Developers interested in working on the HADES codebase. Some experience in R is recommended

OHDSI Medical Imaging Working Group

From pixels to Phenotypes

Enabling observational research with medical imaging data

WG co-leads Seng Chan You and Paul Nagy

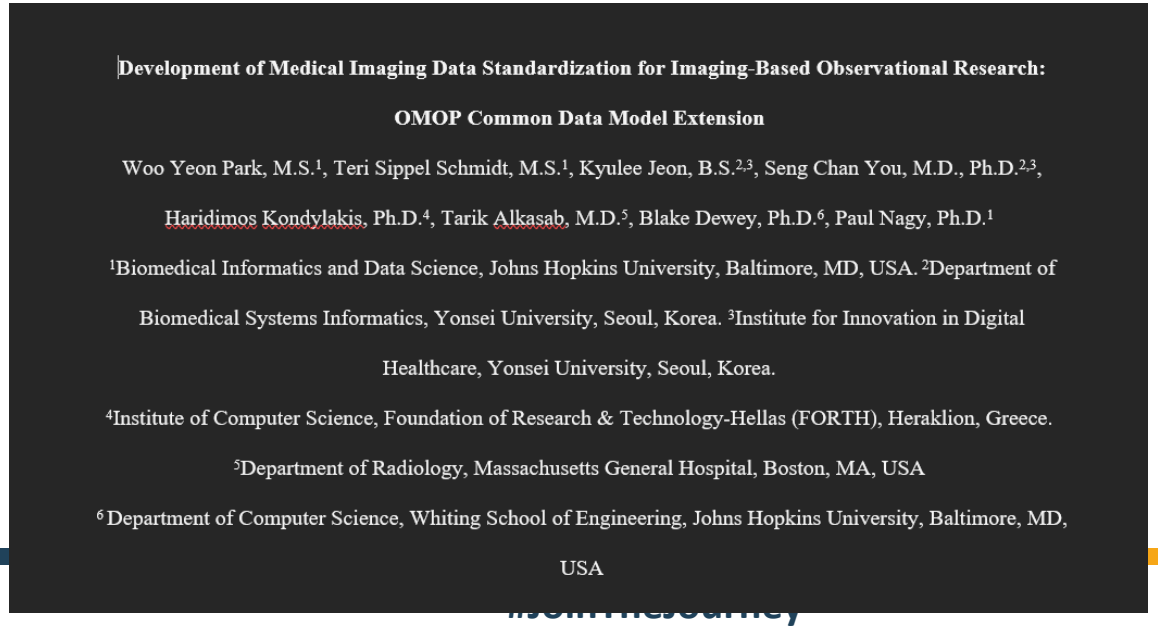
Wednesdays every 2 weeks at 7 AM / 7 PM

Imaging WG Goals

1. Extension to perform cohort definitions in OHDSI for medical imaging research studies.
2. Extension to bring features derived from medical images into the OMOP data model while maintaining provenance.
3. Create refence implementations of infrastructure for reproducible research on medical images.

OKR #1

- Objective: Have CDM group approve the model into the base OMOP model Q3
- Key Results:
  - Publish a draft data model for the imaging extension (Aug 2023)
  - Have Radlex and DICOM vocabularies added to the OMOP vocabulary Q2



Development of Medical Imaging Data Standardization for Imaging-Based Observational Research: OMOP Common Data Model Extension

Woo Yeon Park, M.S.[1], Teri Sippel Schmidt, M.S.[1], Kyulee Jeon, B.S.[2,3], Seng Chan You, M.D., Ph.D.[2,3], Haridimos Kondylakis, Ph.D.[4], Tarik Alkasab, M.D.[5], Blake Dewey, Ph.D.[6], Paul Nagy, Ph.D.[1]

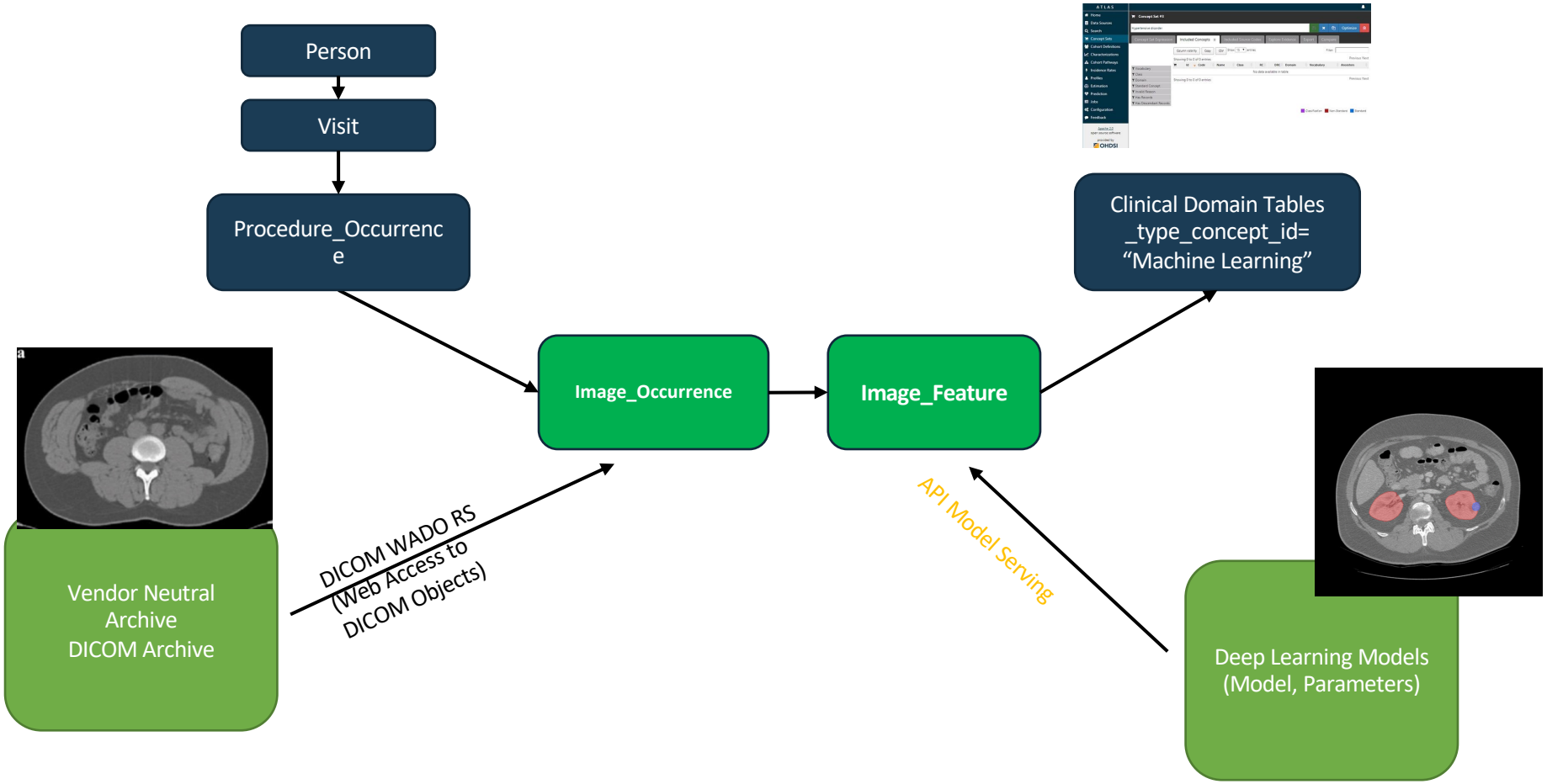
[1]Biomedical Informatics and Data Science, Johns Hopkins University, Baltimore, MD, USA. [2]Department of Biomedical Systems Informatics, Yonsei University, Seoul, Korea. [3]Institute for Innovation in Digital Healthcare, Yonsei University, Seoul, Korea.

[4]Institute of Computer Science, Foundation of Research & Technology-Hellas (FORTH), Heraklion, Greece.

[5]Department of Radiology, Massachusetts General Hospital, Boston, MA, USA

[6] Department of Computer Science, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD, USA

Imaging Extension to OMOP

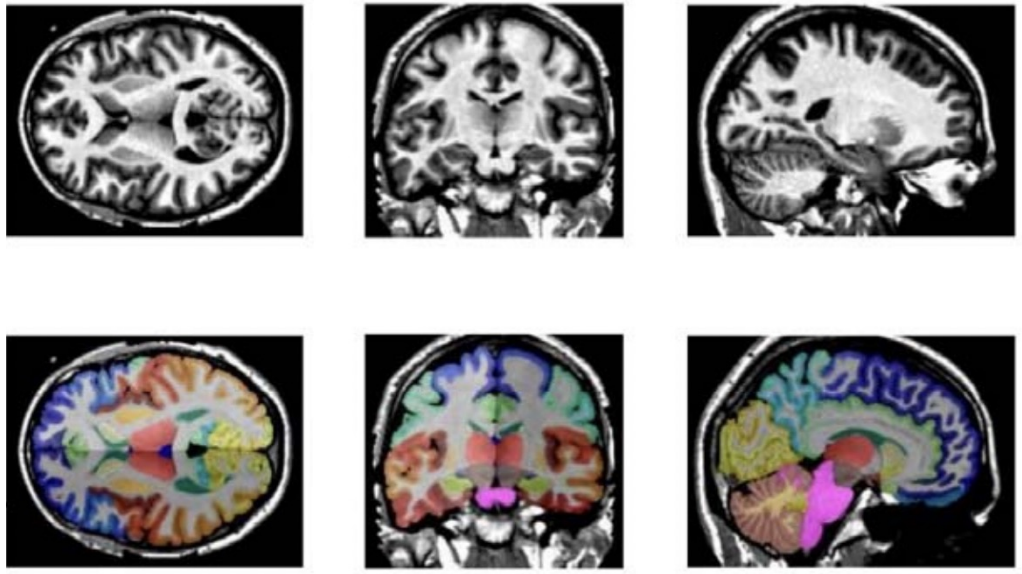


OKR #2

- Objective: Conduct a network study based on the imaging extension

*Seed gift from Gates Ventures in support of Alzheimer's research

MONAI Model Zoo
Wholebrainseg large unest segmentation



March 2024 – 1st imaging data characterization network study
July 2024 – 1st Federated acyclic learning model

OHDSI Vaccine Vocabulary Workgroup 2023

Co-chairs: Yongqun "Oliver" He (UMich), Asiyah Yu Lin (NIH/Axle)

More members: Umich: Yuanyi Pan; UTHealth: Warren Manuel, Rashmie Abeysinghe, Xubing Hao, Licong Cui; Odysseus: Alexander Davydov; IQVIA: Qi Yang

Overall Goal: Harmonize OMOP vaccine vocabularies (e.g.. CVX, Rx-Norm, Rx-Norm extension, CPT4, and HCPCS) using the **Vaccine Ontology (VO)**.

Workgroup Objective and Key Results (OKR) for 2023:

- Objective 1: Build up a **consensus model** of vaccines for OHDSI needs.
  - *Key result*: Basically done.

- Objective 2: **Map** different vaccine representations using the consensus model.
  - *Key results*: (i) Mapped all **CVX** vaccine terms manually and automatically;
    (ii) Developed an **ML method** for automatic mapping.
    (iii) Working and testing on **RxNorm** now.

- Objective 3: **Incorporate** and **evaluate** the VO mapping results to OMOP vocabulary.
  - *Key result*: To do.

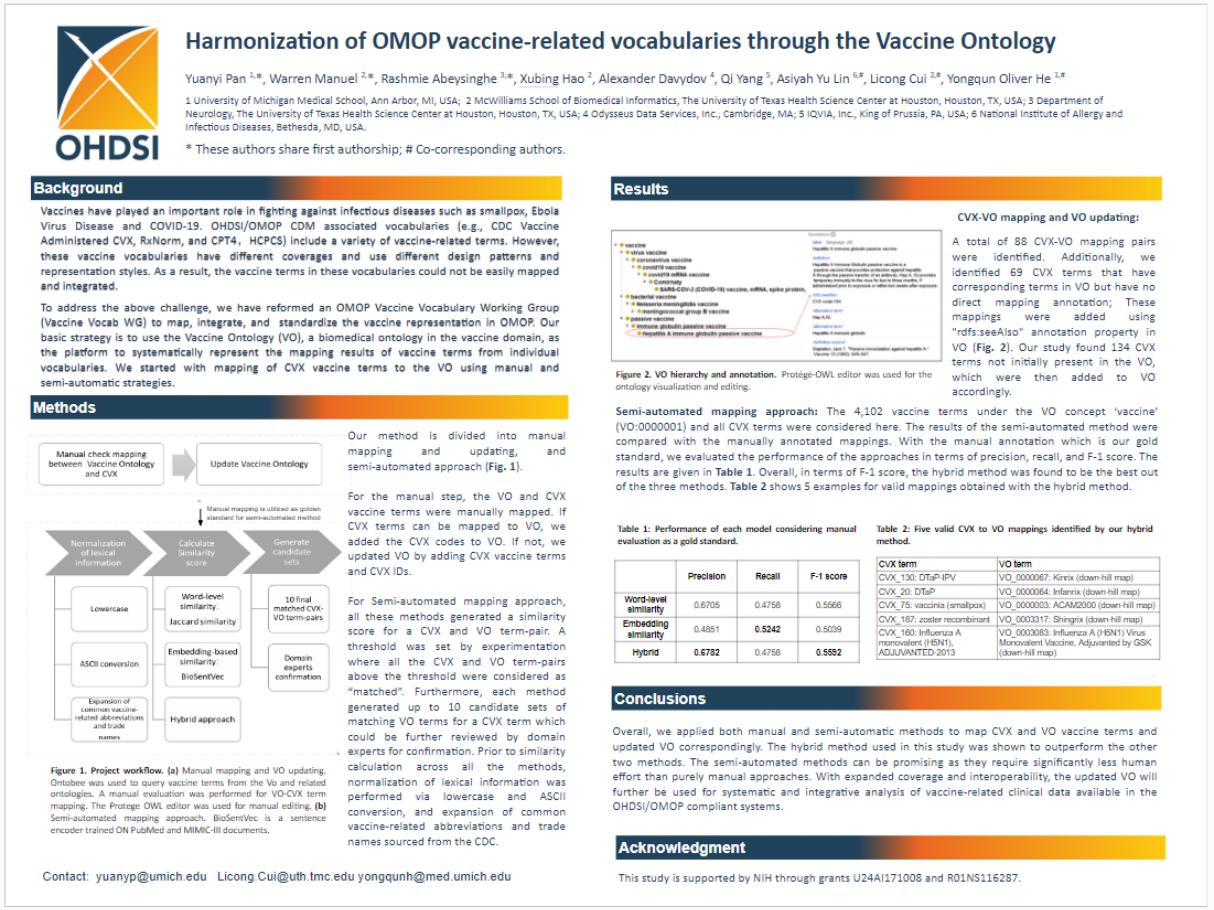
• To present in OHDSI Symposium 2023

Title: Harmonization of OMOP vaccine-related vocabularies through the Vaccine Ontology

Authors: Yuanyi “Penny” Pan*, Warren Manuel*, Rashmie Abeysinghe*, Xubing Hao, Alexander Davydov, Qi Yang, Asiyah Yu Lin#, Licong Cui#, Yongqun Oliver He#

A glimpse of our poster:

(*Note:* Penny will present, and Oliver will attend.)



Harmonization of OMOP vaccine-related vocabularies through the Vaccine Ontology

Yuanyi Pan [1,*], Warren Manuel [2,*], Rashmie Abeysinghe [3,*], Xubing Hao [7], Alexander Davydov [4], Qi Yang [5], Asiyah Yu Lin [6,#], Licong Cui [2,#], Yongqun Oliver He [1,#]

1 University of Michigan Medical School, Ann Arbor, MI, USA; 2 McWilliams School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, USA; 3 Department of Neurology, The University of Texas Health Science Center at Houston, Houston, TX, USA; 4 Odysseus Data Services, Inc., Cambridge, MA; 5 IQVIA, Inc., King of Prussia, PA, USA; 6 National Institute of Allergy and Infectious Diseases, Bethesda, MD, USA.

* These authors share first authorship; # Co-corresponding authors.

Background

Vaccines have played an important role in fighting against infectious diseases such as smallpox, Ebola Virus Disease and COVID-19. OHDSI/OMOP CDM associated vocabularies (e.g., CDC Vaccine Administered CVX, RxNorm, and CPT4, HCPCS) include a variety of vaccine-related terms. However, these vaccine vocabularies have different coverages and use different design patterns and representation styles. As a result, the vaccine terms in these vocabularies could not be easily mapped and integrated.

To address the above challenge, we have reformed an OMOP Vaccine Vocabulary Working Group (Vaccine Vocab WG) to map, integrate, and standardize the vaccine representation in OMOP. Our basic strategy is to use the Vaccine Ontology (VO), a biomedical ontology in the vaccine domain, as the platform to systematically represent the mapping results of vaccine terms from individual vocabularies. We started with mapping of CVX vaccine terms to the VO using manual and semi-automatic strategies.

Methods

Our method is divided into manual mapping and updating, and semi-automated approach (**Fig. 1**).

For the manual step, the VO and CVX vaccine terms were manually mapped. If CVX terms can be mapped to VO, we added the CVX codes to VO. If not, we updated VO by adding CVX vaccine terms and CVX IDs.

For Semi-automated mapping approach, all these methods generated a similarity score for a CVX and VO term-pair. A threshold was set by experimentation where all the CVX and VO term-pairs above the threshold were considered as “matched”. Furthermore, each method generated up to 10 candidate sets of matching VO terms for a CVX term which could be further reviewed by domain experts for confirmation. Prior to similarity calculation across all the methods, normalization of lexical information was performed via lowercase and ASCII conversion, and expansion of common vaccine-related abbreviations and trade names sourced from the CDC.

Figure 1. Project workflow. (a) Manual mapping and VO updating. Ontobee was used to query vaccine terms from the Vo and related ontologies. A manual evaluation was performed for VO-CVX term mapping. The Protege OWL editor was used for manual editing. **(b)** Semi-automated mapping approach. BioSentVec is a sentence encoder trained ON PubMed and MIMIC-III documents.

Results

CVX-VO mapping and VO updating:

A total of 88 CVX-VO mapping pairs were identified. Additionally, we identified 69 CVX terms that have corresponding terms in VO but have no direct mapping annotation; These mappings were added using “rdfs:seeAlso” annotation property in VO (**Fig. 2**). Our study found 134 CVX terms not initially present in the VO, which were then added to VO accordingly.

Figure 2. VO hierarchy and annotation. Protégé-OWL editor was used for the ontology visualization and editing.

Semi-automated mapping approach: The 4,102 vaccine terms under the VO concept ‘vaccine’ (VO:0000001) and all CVX terms were considered here. The results of the semi-automated method were compared with the manually annotated mappings. With the manual annotation which is our gold standard, we evaluated the performance of the approaches in terms of precision, recall, and F-1 score. The results are given in **Table 1**. Overall, in terms of F-1 score, the hybrid method was found to be the best out of the three methods. **Table 2** shows 5 examples for valid mappings obtained with the hybrid method.

Table 1: Performance of each model considering manual evaluation as a gold standard.

	Precision	Recall	F-1 score
Word-level similarity	0.6705	0.4758	0.5566
Embedding similarity	0.4851	**0.5242**	0.5039
Hybrid	**0.6782**	0.4758	**0.5592**

Table 2: Five valid CVX to VO mappings identified by our hybrid method.

CVX term	VO term
CVX_130: DTaP-IPV	VO_0000067: Kinrix (down-hill map)
CVX_20: DTaP	VO_0000064: Infanrix (down-hill map)
CVX_75: vaccinia (smallpox)	VO_0000003: ACAM2000 (down-hill map)
CVX_187: zoster recombinant	VO_0003317: Shingrix (down-hill map)
CVX_160: Influenza A monovalent (H5N1), ADJUVANTED-2013	VO_0003083: Influenza A (H5N1) Virus Monovalent Vaccine, Adjuvanted by GSK (down-hill map)

Conclusions

Overall, we applied both manual and semi-automatic methods to map CVX and VO vaccine terms and updated VO correspondingly. The hybrid method used in this study was shown to outperform the other two methods. The semi-automated methods can be promising as they require significantly less human effort than purely manual approaches. With expanded coverage and interoperability, the updated VO will further be used for systematic and integrative analysis of vaccine-related clinical data available in the OHDSI/OMOP compliant systems.

Acknowledgment

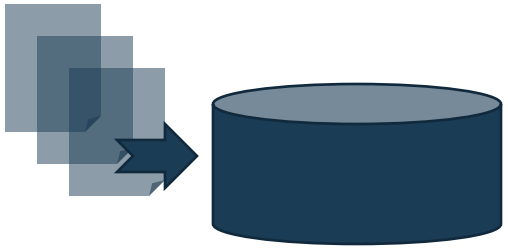
This study is supported by NIH through grants U24AI171008 and R01NS116287.

Contact: yuanyp@umich.edu Licong.Cui@uth.tmc.edu yongqunh@med.umich.edu

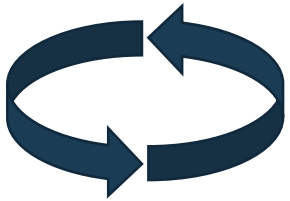
Welcome to our poster!

OHDSI Standardized Vocabularies WG

Workgroup Purpose



- To coordinate the activities of the Vocabulary Team and the OHDSI community to a) provide well curated ontologies in the OHDSI Standardized Vocabularies and b) support transparent pipelines for contributing local content.



- To research and develop approaches and tools to improve quality, transparency and robustness of the OHDSI Vocabularies.

- To connect Workgroups and individuals to enable collaboration around shared ontology needs as well as around tools and methods for Vocabularies management/use in research and ETL.

2023 Objectives

& Key Results

1. Run review sessions for pending changes in vocabularies and their build process

Covered: August 23 release, LOINC-SNOMED hierarchy, MedDRA Overhaul, tracking vocabulary changes

Upcoming: mapping knowledge base, Vocabulary QA/QC and release notes

2. Provide guidance and instructions for community members' contributions

Published and advertised instructions for community contribution (community calls – global and APAC, Vocab WG calls, forums and email distribution)

Soliciting community contribution through stewardship (4 individual and group stewards so far)

Collaborating with the community on more robust pipelines

3. Implement a standard approach for collecting and storing meta-data

SSSOM-compatible meta-data being collected for community contribution

Same standards proposed for Usagi

Implementing the pipeline for MedDRA

HADES
HEALTH ANALYTICS DATA-TO-EVIDENCE SUITE

2023 OKRs

- Regular HADES-wide releases
  - Every 6 months
- More user involvement
  - (not much progress here)
- Have roadmaps, design specifications
  - (not much progress here)
- Improved stability
  - We now have testing servers for all supported database platforms (had to deprecate some platforms for lack of testing server)

Also: Continued Strategus development

Methods Research

2023 goals:

- Create awareness of who is researching what methods in OHDSI / using OMOP
- A sounding board for methods research in progress
- Identify new methods research questions
- Find collaborators

Methods research

2023 methods research highlights:

- Large-scale survival model computations (GPUs)
- Bayesian models adjusting for multiple looks and systematic error (negative controls)
- Likelihood approximations for 1-shot evidence synthesis
- Performance of confounding adjustment when sample size is small
- A better balance metric

2023 updates: OHDSI Medical Device WG

Asiyah Lin

10/9/2023

2023 OKR

Objective 1: Expand the leadership team and establish collaborations across OHDSI and beyond

Key results:

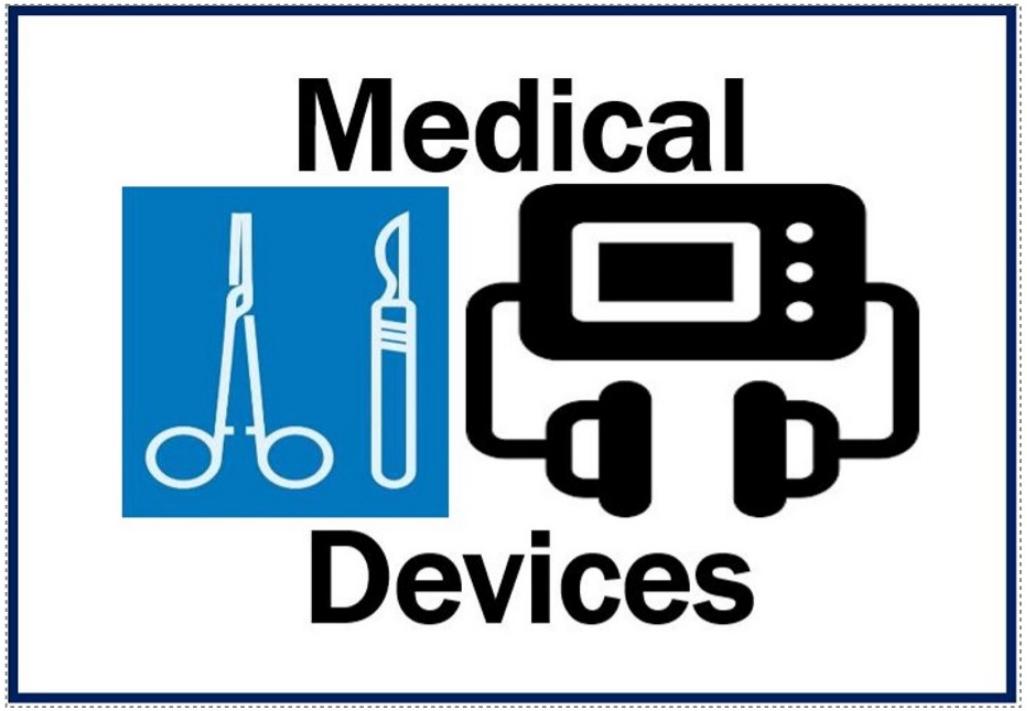
1. Increased meeting frequency: every 3 weeks on Thursdays alternatively 9AM or 12PM Eastern Time.
2. Welcome **Michael Matheny** to join the leadership team!
3. 1Q2023 : Responded to FDA medical device active surveillance RFI by Mar. 30, 2023. - OHDSI WG as a whole.
4. 2-3Q2023: Group members (separately) responded to the NEST/FDA medical device active surveillance RFP:
   - Asiyah Lin led Axle Informatics team;
   - Michael Matheny led VA and Vanderbilt team...
   - Both collaborate with OHDSI.

The first WG OHDSI Symposium Activity!

Sunday, Oct. 22nd, 2023. 1PM – 5PM EDT. @ Campbell room

1. Prototype Device table for CDM
   – Starting point: current device exposure table in the OMOP v5.4,
   and the FHIR DeviceDefinition reource as a starting point.
2. Brainstorming for the year 2024 activities.

Looking for leaders!



Healthcare System Interest Group update

Fall 2023

HSIG meeting at the Symposium

- Not too late to sign up
- Sunday, Oct. 21st 1pm - 5pm
- Guest speakers include Paul Nagy & Kristin Kostka
- Perfect for those thinking about joining the journey, have started their journey, or want more from their journey
- All are welcome!

Themis working group update

Fall 2023

OKRs have been met for the year!

1. To re-establish the Themis WG; 1st & 3rd Thursday of the month @ 9:30am Eastern Time
2. Publish guidelines for the creation, nomination and adjudication process for Themis convention approval; check the Themis GitHub
3. Complete some Themis work; Ratified: Guidance for Drug Exposure days_supply field and ETL guidance for Cohort table

Africa Chapter 2023 Objectives and Key Results

Objective 1: Grow the Africa Chapter to >25 members

- Key Result: 117 members, 13+ African countries, 14+ ex-Africa countries

Objective 2: Prepare promotional video and value propositions

- Key Results

✓ Two value proposition documents: one for researchers (3 pp), another for Ministries and Data custodians (2 pp)

✓ Informational poster for hanging at meetings – one pptx

✓ 4-min video for OHDSI Workgroup page

Africa Chapter 2023 Objectives and Key Results

Objective 3: Develop a pipeline from one or more regional sources into OMOP including data quality assessment

Key Results

- ✓ Rwanda – LAISDAR 15 hospital network coordinated by Rwanda Biomed Center (RBC) and Ghent University
- ✓ Kenya, Malawi – INSPIRE coordinated by African Population Health Research Center (APHRC)
- ✓ Other sites interested but need funding

Objective 4: Identify and pursue funding to support Chapter Activities

Key Results

- ✓ Wellcome Trust ➛ APRHC "Data Science without Borders" Kenya, Cameroon, Senegal, Ethiopia, UK, France Data harmonization to OMOP CDM, Machine Learning etc.
- ✓Others have submitted to open grant calls, pending review or being prepared for government initiatives or philanthropic organizations

Africa Chapter 2023 Objectives and Key Results

Objective 5: Present Africa Chapter at related conferences focused on Africa

- Key Results

✓ 2 presentations at DS-I Africa Virtual Network Exchanges coordinated by Univ. Cape Town

✓ Successful meeting with Africa Chapter members & USAID, CDC and Africa CDC

✓ Upcoming Africa conferences – members offering to display OHDSI Africa Chapter poster

✓OHDSI Global Symposium Africa Chapter Workgroup Activity - Sun Oct 22, 1-5 PM

- Converting maternal hospital admission survey Questions & Answers to OMOP CDM tables, domains, classes, and standard vocabulary concept ids
- Maternal and neonatal health, mapping questionnaires, unique vocabulary
- Teams link for remote participation posted in the Africa Chapter Teams Channel

Africa Chapter Future Objectives

- Build a data source catalog
- Assess vocabulary needs for use cases in the region
- Hold an-in person workshop/conference in Africa (2025)

GIS – Geographic Information Systems

Objectives and Key Results (OKR)

Kyle Zollo-Venecek

Oncology

Objectives and Key Results (OKR)

Asieh Golozar

Surgery & Perioperative Medicine

Objectives and Key Results (OKR)

Evan Minty

Psychiatry

Objectives and Key Results (OKR)

Dmitry Dymshyts

Phenotype Development and Evaluation Workgroup
2023 Update

2023 | 2024

Objective 1: Harden Phenotype Development and Evaluation framework

- ✓ Enable the community to complete 10 phenotypes via the current phenotype development and evaluation process using activities like Phenotype Phebruary (Timeline: 1Q 2023)
- ! Through scientific debate address at least 4 topics of community interest and drive community consensus (Timeline: 2Q 2023)
- ✗ Clarify terminology and scientific definitions and deliver a document that organizes such ideas (Timeline: 2Q 2023)
- ✗ Integrate probabilistic phenotyping into the OHDSI PL (Timeline: 3Q 2023)
- ✓ Write two scientific papers on phenotyping development and evaluation (Timeline: 4Q 2023)

Objective 2: Improve collaboration by enabling community wide participation on Phenotype Development and Evaluation

- ✓ In Phenotype Phebrurary have 10 Phenotypes completed and published in OHDSI Phenotype library as peer reviewed Cohort Definitions (Timeline: 1Q 2023)
- ✓ Promote clinically trained scientists by having at least 5 new clinician collaborators actively engage in Phenotype Phebruary (Timeline: 1Q 2023)

Objective 3: Promote the usage of OHDSI Phenotype library

- ! Complete peer review for 10 (phenotype phebruary) + 10 (additional) phenotypes added to the OHDSI Phenotype Library. (Timeline: 4Q 2023)
- ✓ Formalize submission process and perform at least 2 communication sessions on OHDSI Phenotype Library. (Timeline: 2Q 2023)
- ✓ Execute at least two OHDSI Studies that uses the Cohort Definitions in the OHDSI Phenotype library to generate characterization evidence (Timeline: 3Q 2023)
- ✓ Write a paper on OHDSI Phenotype Library (Timeline: 4Q 2023)

2024 markers: ! ! ! ! ! !